Naturalis Repository

Remotely sensed temperature andprecipitation data improve speciesdistribution modelling in the tropics

V. Deblauwe, V. Droissart, R. Bose, B. Sonké, A. Blach-Overgaard, J.-C. Svenning, J. J. Wieringa, B. R. Ramesh, T. Stévart and T. L. P. Couvreur

Downloaded from
https://doi.org/10.1111/geb.12426

Article 25fa Dutch Copyright Act (DCA) - End User Rights
This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with consent from the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available following a reasonable period after the work was first published, provided that reference is made to the source of the first publication of the work.

This publication is distributed under the Naturalis Biodiversity Center 'Taverne implementation' programme. In this programme, research output of Naturalis researchers and collection managers that complies with the legal requirements of Article 25fa of the Dutch Copyright Act is distributed online and free of barriers in the Naturalis institutional repository. Research output is distributed six months after its first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and copyrights owner(s) of this work. Any use of the publication other than authorized under this license or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the department of Collection Information know, stating your reasons. In case of a legitimate complaint, Collection Information will make the material inaccessible. Please contact us through email: collectie.informatie@naturalis.nl. We will contact you as soon as possible.

Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2016) **25,** 443–454

Global Ecology and Biogeography A Journal of Macroecology

RESEARCH PAPER

Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics

V. Deblauwe[1,2,3]*, V. Droissart[2,3,4,5], R. Bose[4,6], B. Sonké[3], A. Blach-Overgaard[7], J.-C. Svenning[7], J. J. Wieringa[8], B. R. Ramesh[6], T. Stévart[2,5] and T. L. P. Couvreur[1,3,8]

[1]*Institut de Recherche pour le Développement (IRD), UMR DIADE, F-34394 Montpellier, France,* [2]*Herbarium et Bibliothèque de Botanique Africaine, Université Libre de Bruxelles, B-1050 Brussels, Belgium,* [3]*Laboratoire de Botanique Systématique et d'Écologie, Département des Sciences Biologiques, École Normale Supérieure, Université de Yaoundé I, Yaoundé, Cameroon,* [4]*Institut de Recherche pour le Développement (IRD), UMR AMAP, F-34398 Montpellier, France,* [5]*Missouri Botanical Garden, Africa and Madagascar Department, 63166-0299 St Louis, MO, USA,* [6]*French Institute of Pondicherry, Department of Ecology, Puducherry 605001, India,* [7]*Section for Ecoinformatics and Biodiversity, Dept of Bioscience, Aarhus Univ., DK-8000 Aarhus, Denmark,* [8]*Naturalis Biodiversity Centre, Botany Section, 2333 CR Leiden, The Netherlands*

ABSTRACT

Aim Species distribution modelling typically relies completely or partially on climatic variables as predictors, overlooking the fact that these are themselves predictions with associated uncertainties. This is particularly critical when such predictors are interpolated between sparse station data, such as in the tropics. The goal of this study is to provide a new set of satellite-based climatic predictor data and to evaluate its potential to improve modelled species–climate associations and transferability to novel geographical regions.

Location Rain forests areas of Central Africa, the Western Ghats of India and South America.

Methods We compared models calibrated on the widely used WorldClim station-interpolated climatic data with models where either temperature or precipitation data from WorldClim were replaced by data from CRU, MODIS, TRMM and CHIRPS. Each predictor set was used to model 451 plant species distributions. To test for chance associations, we devised a null model with which to compare the accuracy metric obtained for every species.

Results Fewer than half of the studied rain forest species distributions matched the climatic pattern better than did random distributions. The inclusion of MODIS temperature and CHIRPS precipitation estimates derived from remote sensing each allowed for a better than random fit for respectively 40% and 22% more species than models calibrated on WorldClim. Furthermore, their inclusion was positively related to a better transferability of models to novel regions.

Main conclusions We provide a newly assembled dataset of ecologically meaningful variables derived from MODIS and CHIRPS for download, and provide a basis for choosing among the plethora of available climate datasets. We emphasize the need to consider the method used in the production of climate data when working on a region with sparse meteorological station data. In this context, remote sensing data should be the preferred choice, particularly when model transferability to novel climates or inferences on causality are invoked.

*Corresponding author: Vincent Deblauwe, Institut de Recherche pour le Développement (IRD), UMR DIADE, F-34394 Montpellier, France.
E-mail: vincent.deblauwe@ulb.ac.be

Keywords

Association test, CHIRPS, ecological niche model, GLM, habitat suitability, MaxEnt, MODIS, null model, TRMM, WorldClim.

INTRODUCTION

Species distribution models (SDMs) are playing an increasingly important role in many fields of biology (Peterson, 2011a). To reach the highest possible model generality, SDMs are usually calibrated either partially or completely on climatic variables, given their direct influence on species distributions (Guisan & Zimmermann, 2000). Several sources of uncertainty in the SDM

DOI: 10.1111/geb.12426
http://wileyonlinelibrary.com/journal/geb

output have been identified, including modelling strategies (Elith *et al*., 2006; Dormann, 2007), the omission of causal predictors (Bedia *et al*., 2013; Barbet-Massin & Jetz, 2014), species ecology and the quality of occurrence data (Hopkins, 2007; Dormann *et al*., 2008; Dobrowski *et al*., 2011) and the assumption of niche conservatism through space and time (Peterson, 2011b). However, uncertainties arising from the quality of the climatic datasets used, which are themselves predicted values of current climate, have received little attention (but see Soria-Auza *et al*., 2010; Bedia *et al*., 2013; Fernandez *et al*., 2013).

The most popular source of temperature and precipitation data among ecologists is the freely accessible WorldClim global dataset (Hijmans *et al*., 2005). The developers of the dataset aimed to improve earlier datasets that employed the same methodology, such as the Climatic Research Unit (CRU) high-resolution climate data (New *et al*., 1999, 2002), by producing the data at a higher spatial resolution (down to 1 km) and by the inclusion of improved elevation and station datasets.

At the local scale (< 50 km), climate patterns are mostly affected by topographical features and the presence of water bodies (Daly, 2006). Locally sharp climatic features such as rain shadows cannot be adequately represented unless: (1) a predictor that is a proxy for this influence is incorporated in the interpolation, or (2) there are a sufficient number of stations to capture this dependence as a function of the considered predictors. Indeed, using the WorldClim dataset, Fernandez *et al*. (2013) found that SDM uncertainty was related to topographic heterogeneity, inter-annual variability and distance to the closest weather station.

In contrast to developed countries, most tropical areas suffer from a low density of weather stations (Washington *et al*., 2013) and intermittent observations, leading to inconsistent predictions of climate patterns (New *et al*., 2002; Fernandez *et al*., 2013) across datasets, even at large scales (Fig. 1; and see cross-validation in Hijmans *et al*., 2005). However, it is notably in the tropics that the need for freely available global data is important, given the difficulty of obtaining digital maps of regional monthly climatology. Indeed, over a third of SDM studies related to the tropics used global dataset, whereas to less than a quarter of studies focusing outside the tropics did (Fig. S1 in Supporting Information).

In recent years, the development of high-resolution global satellite data has permitted the publication of global gridded

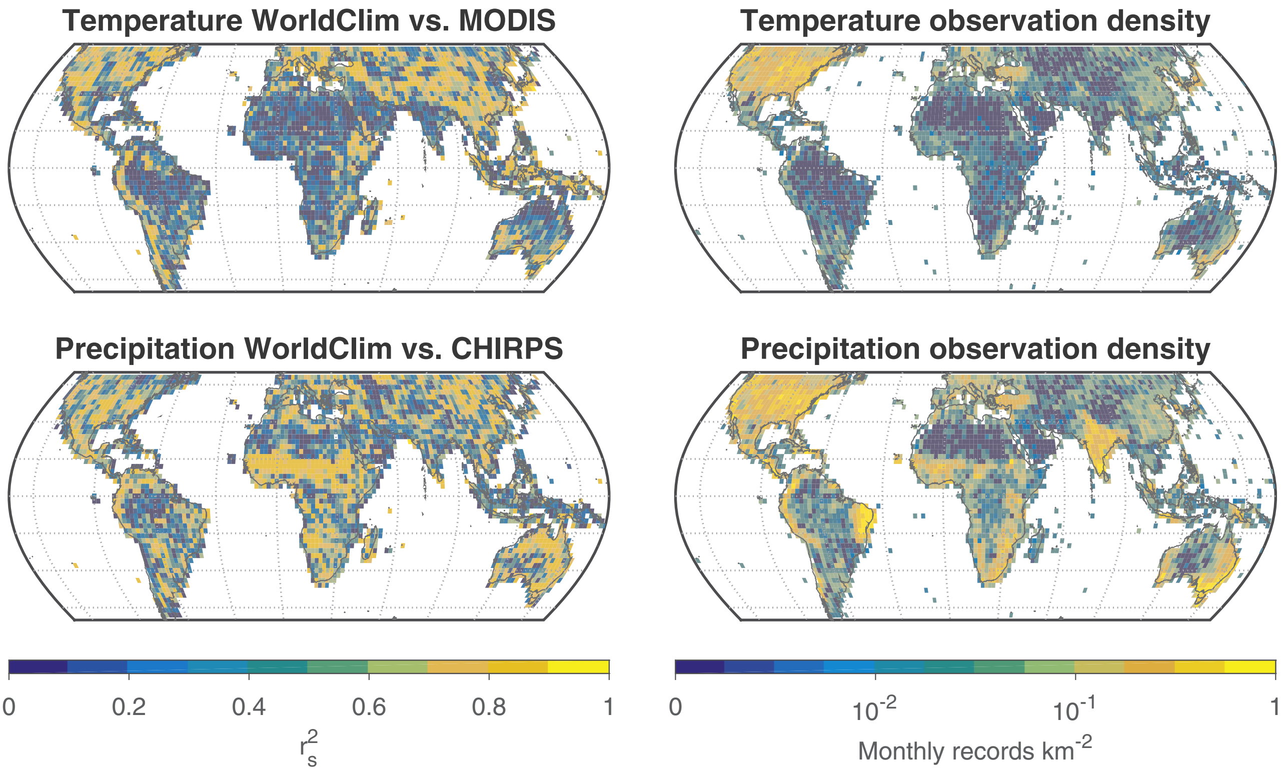

Figure 1 The global pattern of climate dataset uncertainties (left panel) matches that of weather station data density (right panel). Areas with sparse station data, as is often the case in tropical rain forests, are also areas of greater inconsistency between the remote sensing and WorldClim climate patterns (low values in darker shades or blue in the online version). The colour/grey scale represents values of Spearman's rank correlation coefficients (r_s^2) between annual mean temperature from WorldClim and MODIS (top left) and annual precipitation from WorldClim and CHIRPS (bottom left), and density of monthly observations of temperature (top right) and precipitation (bottom right). All values were calculated from non-overlapping windows of 2° × 2°. Station data (1950–2000) were obtained from the Global Historical Climatology Network (Vose *et al*., 1992), and the World-wide Agroclimatic Database from the Food and Agriculture Organization (FAO, 2001), the main data sources for WorldClim.

datasets describing a variety of physical properties of the atmosphere and land surface. The use of remote sensing data as predictors in SDM is not new (see Bradley *et al.*, 2012). However, these studies usually rely on the spectral signature of vegetation rather than on surface or biophysical properties having a direct influence on species survival and reproduction success.

Conversely, space-borne estimates of climatic variables do not measure vegetation directly and are therefore good candidates for characterizing species climatic preferences (Bradley *et al.*, 2012). They are potentially superior to ground station interpolations because: (1) they are a result of spatially continuous direct sampling from space of climatic proxies and hence are not derived from punctual records on the ground, and (2) they often offer high sampling rates in space and time through several decades. Interestingly, to our knowledge, such datasets have rarely been used and never been assessed for SDMs.

In this study we compared SDMs built using a combination of WorldClim and CRU data and the three following (near-) global-scale and publicly available remote-sensing derived datasets: Tropical Rainfall Measuring Mission (TRMM, NASA/JAXA) (Kummerow *et al.*, 1998), Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) (Funk *et al.*, 2015) and NASA Moderate Resolution Imaging Spectroradiometer land surface temperature (MODIS) (Wan & Dozier, 1996; Wan, 2014). We tested the hypothesis that remotely sensed climate datasets improve SDM accuracy in rain forest areas where weather stations are sparse. Most climate datasets are based on station data, making it difficult to assemble a global set of independent observations to gauge the accuracy of these datasets. Our study thus also constitutes an assessment of climate predictions using the entirely independent information contained in species distributions.

We explored two complementary aspects of model quality using methods that take into account the inflation of accuracy metrics due to the spatial structure present in both species distributions and climatic variables (Beale *et al.*, 2008; Chapman, 2010; Deblauwe *et al.*, 2012; Heikkinen *et al.*, 2012; Hijmans, 2012; Rapacciuolo *et al.*, 2012): (1) the model's ability to express a true association between climate and species distribution, and (2) model transferability (or generalization), i.e. the ability of the model to make accurate predictions when transferring the above association onto novel data. For the first, we devised an original null model to determine the expected distribution of the goodness-of-fit statistic under the null hypothesis of no association between the observed species distribution and climate patterns.

METHODS

Climatic data sets

The main characteristics of the three temperature (WorldClim, CRU and MODIS) and four precipitation (WorldClim, CRU, TRMM and CHIRPS) datasets included in this study are summarized in Table 1 and are detailed in Appendix S1.

Since the native ground resolution of data varies between sources, we resampled each monthly grid to a resolution of 0.1° × 0.1° (i.e. 6′). This was deemed to be an optimal operational scale as it permits coarser grid modelling while maintaining most of the accuracy of the finer scales (Seo *et al.*, 2009). At a higher spatial resolution, the interpolation of climate is challenging, and may not be accurate in many regions of the world (Daly, 2006). On the other hand, very few ecological studies will consider the coarsest resolution (i.e. 0.25° for TRMM and 10′ for CRU) which would not represent many of the important variations in climate and ecological processes (Seo *et al.*, 2009). The aggregation of WorldClim, CHIRPS and MODIS grids was achieved through non-overlapping moving window averaging. The MODIS values severely contaminated by clouds and heavy aerosols were discarded when averaging values through space and time. Bilinear interpolation was applied to CRU and TRMM grids to match the 0.1° cell-size.

As they are a popular reference among ecologists, WorldClim data were chosen as a baseline against which to compare every other source of temperature (T) or precipitation (P) data. We constructed five hybrid temperature–rainfall datasets which included WorldClim data for either temperature (T^{WC}) or precipitation (P^{WC}). These are hereafter named after the source of data as follows: T^{MODIS} P^{WC}, T^{CRU} P^{WC}, T^{WC} P^{CHIRPS}, T^{WC} P^{TRMM} and T^{WC} P^{CRU}. From the average monthly

Table 1 Description of the climate dataset used in this study.

Dataset	Native resolution	Type of data	Period of record used in this study	Method	URL	Reference
WorldClim	0.5′, 2.5′, 5′, 10′	T, P	1950–2000	Splines interpolation between stations	http://www.worldclim.org	Hijmans *et al.* (2005)
CRU CL v.2.0	10′	T, P	1961–90	Splines interpolation between stations	http://www.cru.uea.ac.uk	New *et al.* (2002)
TRMM 3B43 v.7	15′	P	1998–2013	Remote sensing	http://disc.sci.gsfc.nasa.gov/	Kummerow *et al.* (1998)
CHIRPS v.2.0	3′	P	1981–2013	Compilation (reanalysis)	http://chg.geog.ucsb.edu/	Funk *et al.* (2015)
MODIS LST MOD11C3 v.5	3′	T	2001–13	Temperature derived from thermal infrared band	https://lpdaac.usgs.gov/	Wan & Dozier (1996)

T, temperature; P, precipitation.

precipitation and/or maximum and minimum temperature values of each hybrid climate dataset, we derived the climatic variables described in the ANUCLIM scheme (Xu & Hutchinson, 2011) to generate a common set of variables that are more biologically meaningful than raw monthly means. These variables are very popular among SDM researchers as they represent annual trends, seasonality and extreme or limiting environmental factors. For consistency with WorldClim's adaptation of ANUCLIM, the standard deviation was used for temperature seasonality. To avoid introducing any dependence between temperature and precipitation data we excluded variables derived from a combination of both (see Table S1).

Because transferability is affected by overfitting, which is likely to be more serious in high-dimensional environmental spaces, we used principal components analysis (PCA) to create synthetic predictor axes that summarized the variation in climatic parameters in fewer dimensions. For each study region (see below), we retained the components that accounted for ≥ 98% of the variation.

Species occurrences and study regions

We focused on species endemic to three different tropical regions: Atlantic Central Africa (ACA), the Western Ghats (WG) and the Amazon Basin (AB). For each region we included all areas classified in the World Wildlife Fund (WWF) terrestrial eco-regions as the 'Tropical and subtropical moist broadleaf forests biome' (simply 'rain forest' hereafter) (Olson *et al.*, 2001) extended by a 1° buffer (maps presented in Figs S2–4). To create the above synthetic climate variables, six components of the PCA in ACA and AB and five in WG were used.

We used high-quality angiosperm plant databases for both ACA and WG and an open access database for AB (see Table S2). For ACA, we used a database of herbarium specimens of four major plant families that has been checked and georeferenced by taxonomical experts. A wide range of biological types are represented in these families: epiphytic or terrestrial herbs (Orchidaceae), understorey shrubs (Rubiaceae) and trees as well as lianas (Annonaceae, Arecaceae, Rubiaceae). For WG, we used the *Atlas of endemics of the Western Ghats* (Ramesh & Pascal, 1997) which reports species occurrences based on herbaria specimens, data published in the literature and results of field surveys conducted by the botanists of the French Institute of Pondicherry. Finally, we included georeferenced occurrences of all angiosperm taxa for AB available from the Global Biodiversity Information Facility (http://www.gbif.org; accessed September 2014). We used the Taxonomic Name Resolution Service v.3.2 (Boyle *et al.*, 2013) to update all taxonomic names according to TROPICOS (Missouri Botanical Garden, USA). Occurrences falling on entire latitude and longitude coordinate values (degrees with no decimals) were deemed possibly inaccurate and removed. Finally, we reduced the strong sampling bias in the Amazonian data (Hopkins, 2007) by sampling at random a maximum of two occurrences per square of 1° latitude × 1° longitude.

For all datasets, we discarded all species having more than 10% of occurrences outside the rain forest biome. For the remaining species, only occurrences within the biome were used. For each species a single occurrence per grid cell was retained, and species with fewer than 10 (ACA and WG dataset) or 30 (AB datasets) unique occurrences were discarded.

To account for variability in species geographical ranges we used a smaller area of analysis species-wise defined as a rectangle extending 3° for ACA and AB and 2° for WG in every cardinal direction outside the geographical range of recorded occurrences (Fig. 2a, e). This approach restricts the study extent to areas accessible to species of interest and minimizes the bias of estimated coefficients and evaluation statistics (Barve *et al.*, 2011).

SDM algorithms

SDMs were calibrated using two algorithms representative of different modelling approaches: (1) MaxEnt (Phillips *et al.*, 2006), a machine-learning modelling technique that can potentially match highly nonlinear complex relationships, and (2) generalized linear modelling (GLM). MaxEnt offers relatively good accuracy when modelling species distributions in the present (Elith *et al.*, 2006) and for projection onto novel environments (Hijmans & Graham, 2006; Heikkinen *et al.*, 2012). We calibrated MaxEnt models using default settings. GLM is a parametric method that is comparable to MaxEnt in terms of extrapolative accuracy (Heikkinen *et al.*, 2012; Rapacciuolo *et al.*, 2012). Our aim here was to evaluate the relative effect of different climate sources and whether this is robust to the choice of algorithm, not absolute model accuracy. Thus, for the sake of computational efficiency and to avoid overfitting models based on a limited set of occurrences, GLMs were calibrated on constant and linear terms only. This simpler GLM specification affected all the models that we compared in the same way.

Species–climate association

Only if a species' distribution is actually dependent on the included variables will a model provides useful predictions. This assumption is difficult to test because standard association statistics are often inflated by the spatial dependence present in both species distributions and environmental gradients (Deblauwe *et al.*, 2012). We therefore designed an original method to generate null geographical distributions of species presence that mimic the distribution pattern of the species but without any association to climate (Fig. 2).

We used an iterative procedure known as the mimetic point process (Pélissier & Goreaud, 2015) to generate random gridded distributions of presences with the same resolution, geographical extent, number of occupied cells and spatial dependence (clustering or dispersion) as the actual observations. The spatial dependence is summarized over a range of distances by the variance-stabilized Ripley *K*-function, called the *L*-function (Ripley, 1976). The algorithm performs a stepwise depletion–replacement of presences minimizing the difference between the actual and simulated *L*-function until a given value of the coefficient of variation of the root mean square error of the difference is achieved. Alternatively, the simulation will stop after a specific number of iterations has been reached (determined by the number of presence points). Our null model therefore randomizes any existing relationship between species distribution and climate while preserving the density and structure of the geographical distribution (Fig. 2c, g). A correction for 'border effect' was not required since very few occurrences were located outside the model's training area. A detailed description of the procedure is available in Goreaud *et al.* (2004). A similar randomization procedure displayed adequate statistical power to detect species–climate dependence (Beale *et al.*, 2008). To reduce computation time, instead of the usual 999 or more simulations we generated 499 Monte Carlo simulations for each species and used GLM to calibrate the corresponding SDMs. The probability of obtaining an area under the receiver operating characteristic curve (ROC-AUC, AUC hereafter) equal or superior to the observed AUC under the null hypothesis, i.e. the *P*-value, was computed as the frequency at which $\mathrm{AUC}_{\mathrm{null}} \geq \mathrm{AUC}_{\mathrm{obs}}$ (single-tailed test) from a total of 500 AUCs.

Geographical partitioning and SDM transferability

To challenge the SDM projections across unsampled areas of geographical space, we separated the occurrence points and study regions for each species into two equal sets based on whether they fell above or below the species' median latitude. This corresponds to a two-fold cross-validation where occurrence points are partitioned into training and test sets according to a geographical criterion rather than at random in order to maximize independence between them. We first created a model using the climatic values at occurrence sites in one study region (region A), then transferred the model by applying it to the climatic values in the second region (B) (Peterson & Shaw, 2003). The background information and pseudo-absences were drawn from the calibration region A. Internal evaluation a model was based on its ability to accurately represent the

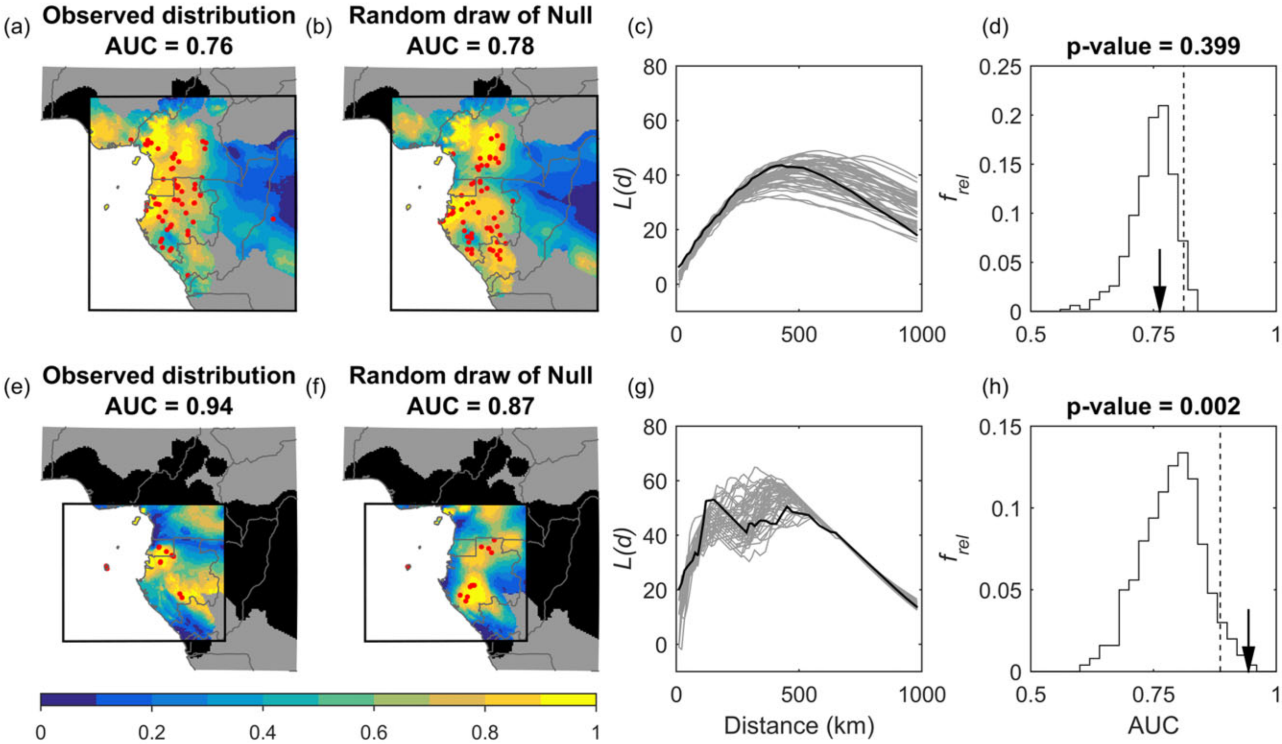

Figure 2 Example of the mimetic point process method for simulating null species distribution models used in this study. Generalized linear model (GLM) predictions for two species in Atlantic Central Africa with contrasting distributions: (a)–(d) *Anonidium mannii* (Oliv.) Engl. & Diels with a widespread, relatively homogeneous distribution; (e)–(h) *Rangaeris trilobata* Summerh. with a clumped distribution. The left column (a, e) shows the actual distribution model and associated area under the receiver operating characteristic curve (AUC) value as obtained from GLMs using the WorldClim dataset. The regions outside rain forests are in grey. The black regions correspond to rain forest area outside the training area (delineated by inset rectangle). The second column of graphs (b, f) represents a randomly chosen null model (from among 499 simulations) and its associated AUC value. The third column (c, g) represents the spatial dependence (see Methods) as a function of distance between observed presences (black line) and 50 simulations (grey lines). The histograms on the far right (d, h) represent the distribution of the null AUC values for the 499 null species distributions. On the graph is indicated the one-tailed 95% confidence interval (dashed vertical line), the AUC value for the actual distribution (arrow) and the corresponding *P*-value. The match between WorldClim predictors and distribution for *A. mannii* and *R. trilobata* was found to be, respectively, not significant and significant, yet both are significant when using MODIS temperatures (not shown).

presence and (pseudo-)absence values that were used for calibration in region A. A model's ability to discriminate between occurrences and (pseudo-)absences in region B was considered as an external evaluation. The AUC was used to assess the models' internal ($\mathrm{AUC}_{\mathrm{A}\rightarrow\mathrm{A}}$) and external ($\mathrm{AUC}_{\mathrm{A}\rightarrow\mathrm{B}}$) performances. A second model run was then applied for each species by inversing the training and testing roles of regions A and B. The four resulting AUCs were combined into a single transferability index (TI), introduced by Randin *et al.* (2006):

$$\mathrm{TI}=\frac{1-|\mathrm{AUC}_{\mathrm{A}\rightarrow\mathrm{A}}-\mathrm{AUC}_{\mathrm{A}\rightarrow\mathrm{B}}|-|\mathrm{AUC}_{\mathrm{B}\rightarrow\mathrm{B}}-\mathrm{AUC}_{\mathrm{B}\rightarrow\mathrm{A}}|}{1+2\big||\mathrm{AUC}_{\mathrm{A}\rightarrow\mathrm{A}}-\mathrm{AUC}_{\mathrm{A}\rightarrow\mathrm{B}}|-|\mathrm{AUC}_{\mathrm{B}\rightarrow\mathrm{B}}-\mathrm{AUC}_{\mathrm{B}\rightarrow\mathrm{A}}|\big|}.$$

As the AUC is bounded between 0 (inverse discrimination) and 1 (perfect discrimination), the TI may vary between –1 and 1. However, AUCs do not usually take values below 0.5 (no discrimination) and hence TI ranges between 0 and 1 in most cases.

Statistical analyses

The null model and TI results were compared with the T^{WC} P^{WC} baseline separately for every model setting (algorithm × region × input climate data). The frequency of change in the detection of species as significant or not was assessed using McNemar's chi-square test of marginal frequency. The null hypothesis that TI gain comes from a distribution with zero median was tested using the two-tailed Wilcoxon signed-rank test. Due to the multiplicity of individual hypotheses tested, we computed q-values (the adjusted P-values of Storey, 2003) following Benjamini & Hochberg (1995) to control for the expected proportion of incorrectly rejected null hypotheses among all rejections (false discovery rate, FDR). This was done for the following families of hypotheses: randomization test (451 hypotheses for each of the six types of climate data), randomization test comparison by climate data (15 hypotheses corresponding to five comparisons with the baseline by region of origin) and TI gain comparison (15 hypotheses corresponding to five comparisons with the baseline by region of origin for each algorithm). We chose not to apply a more stringent procedure such as the Holm–Bonferroni method (family-wise) to maintain reasonable statistical power for individual inferences.

A higher accuracy in the description of species preferences should result in higher model transferability (Rödder & Lötters, 2010). We therefore predicted that the variation in q-value, i.e. evidence against the null hypothesis (Murtaugh, 2014) of species–climate independence, observed for the different sources of climate data would be related to an inverse variation in TI.

All data preparation and analyses were performed using our own MATLAB® routines.

RESULTS

Species–climate associations

A total of 451 species were selected for modelling (174, 170 and 107 in ACA, WG and AB, respectively). We fitted GLM to both the true and simulated distributions using WorldClim and the five hybrid climate datasets, resulting in a total of 1,353,000 models. For WorldClim-based models, training AUC scores were above 0.7 for 95% of the species (median 0.87). However, in only 40% of the cases were the AUCs of models of real species distributions significantly higher than the AUCs of simulated distributions at the 0.05 level (Fig. 3a). This proportion varied strongly among the studied regions with 20% in ACA, 75% in WG and 20% in AB (Fig. 3a). When WorldClim temperature data were replaced by MODIS these numbers significantly increased in every instance, with 32% in ACA, 96% in WG and 35% in AB (an overall increase of 40%). Inclusion of CHIRPS precipitation resulted in the second most important difference with WorldClim, with a 22% overall increase driven by substantial differences in ACA and WG. TRMM precipitation led to an overall decrease of 16%, mainly due to results in WG. Incorporating CRU temperature data increased the number of significant species in AB, but did not result in any differences elsewhere, whereas using CRU precipitation data led to a significant decrease in AB only. Choosing alternative significance levels of 0.1 or 0.01 did not affect these tendencies (see Fig. S4 for the cumulative distribution function of q-values).

Model transferability

The geographical transferability was assessed species-wise for each possible combination of algorithm (MaxEnt and GLM) and the six climate datasets, resulting in a total of 10,824 models. Mean training AUC of models using MaxEnt (ranging from 0.9 to 0.91) gave superior values to those obtained with GLM (0.88 to 0.89). While the choice of climate data had little impact on the resulting training AUC (result not shown), it greatly affected the transferability of the models.

Differences in transferability when calibrating models with WorldClim data and with data partly replaced by climate information from other sources revealed contrasting patterns (Fig. 3b). The increase in model transferability (positive gain) as a result of incorporating MODIS temperature was robust to the choice of algorithm and geographical region. Apart from the TI gains using MaxEnt in ACA and AB, which were not significant, replacing WorldClim precipitation data by CHIRPS estimates led also to a significant gain in transferability whereas TRMM precipitation data resulted in a significant loss of transferability (negative gain) when using MaxEnt models for ACA and AB. Despite the fact that using TRMM data in GLM led to an important reduction in the number of WG species deemed significant (see above), this change was not reflected by a loss in TI in any region. Incorporating CRU temperature data resulted in an inconsistent pattern, with a loss of transferability in ACA but a gain in WG and AB. In AB, CRU temperature data were related to a strong increase in significant species–climate association. On the other hand, models using CRU precipitation data were similar in terms of transferability to those using WorldClim values, except for some loss in AB with GLM.

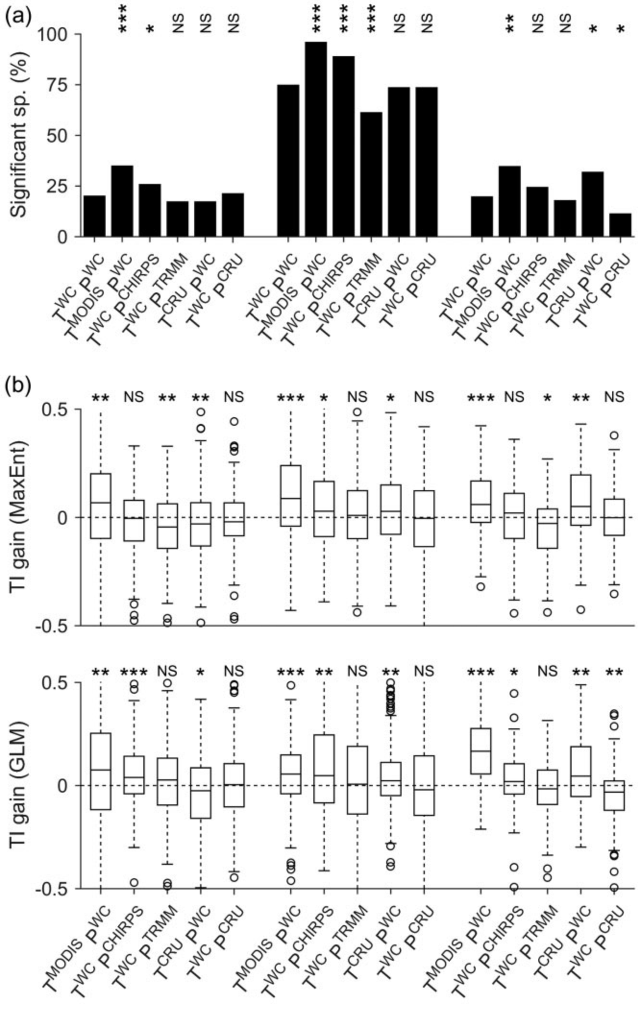

Figure 3 Impact of different climatic datasets on the accuracy of rain forest species distribution models. The three regions are plotted in separate groups from left to right: Atlantic Central Africa, the Western Ghats and the Amazon Basin. Results are shown for every combination of temperature (T) and precipitation (P) data. (a) Proportion of species for which a significant association to climate at level of 0.05 has been detected. Results of the McNemar's chi-square test for individual comparison with WorldClim data (T^{WC} P^{WC}) are indicated above each bar. (b) Box-plot of the species-wise difference between transferability index (TI) of models calibrated on T^{WC} P^{WC} and five hybrid climate models with positive (improvement) and negative (loss) gain of values in TI relative to T^{WC} P^{WC}-based models. The horizontal dashed line corresponds to no difference. Boxes delimit the interquartile range with a solid horizontal line at the median. Whiskers extend to the most extreme data points that are not considered outliers (i.e. 1.5 times interquartile range) and outliers are plotted as circles. Above each box is indicated the result of Wilcoxon signed-rank test for the null hypothesis that TI gain comes from a distribution with zero median. ***q-value ≤ 0.001; ** 0.001 < q-value ≤ 0.01; * 0.01 < q-value ≤ 0.05; NS (not significant), q-value > 0.05.

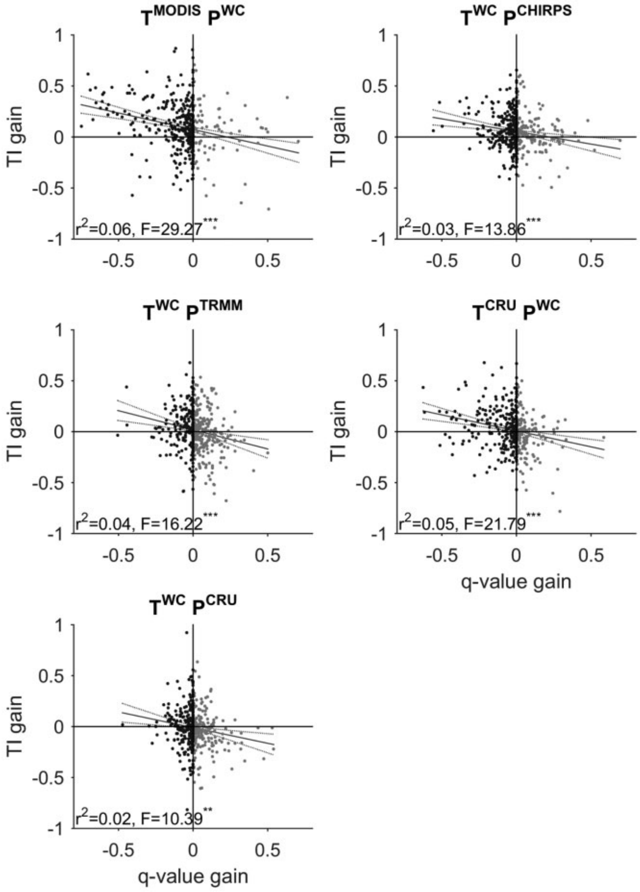

Figure 4 Transferability index (TI) gain as a function of the null model's q-value gain for each climate model. Results are shown for every combination of temperature (T) and precipitation (P) data. Each dot corresponds to a species. Linear regressions are shown with 95% confidence bounds (dashed curves) around the fitted line (solid line). Coefficient of determination, r^2, and the F statistic of the linear fit versus the constant model are indicated along with level of significance: ***P-value ≤ 0.001; **0.001 < P-value ≤ 0.01.

Relationship between randomization test and transferability

The inverse relation between the variation in q-value and the change in model transferability (TI) is clear in our data (Fig. 4). In other words, when we replaced WorldClim temperature or precipitation data by another dataset, the improved description of the species climatic envelope (lower q-value) also results in improved model transferability (higher TI), and vice versa.

DISCUSSION

Species distribution modelling in regions where weather stations are scarce is problematic. The widely used interpolated climate grids relying on spatial and topographic predictors (e.g. WorldClim or CRU) are unlikely to capture complex spatial features of the climate (Daly, 2006) and therefore add their own uncertainty to models (Soria-Auza *et al.*, 2010; Fernandez *et al.*, 2013). Here, we provide a quantitative assessment of: (1) the degree to which interpolated climate grids are accurate enough for species distribution modelling in three tropical rain forest areas, and (2) the improvement from using remote sensing-derived climate data in such areas. Our results show a significant improvement in the modelled association and its transferability when incorporating either MODIS temperature or CHIRPS precipitation data, both datasets being derived from remote sensing data (Fig. 3). The specific gain in transferability of different climate data sources relative to the WorldClim baseline varied slightly between MaxEnt and GLM algorithms, but the trends remained constant, suggesting that our findings are robust to the choice of algorithm.

Species–climate association and transferability

The training AUC scores of SDMs calibrated with WorldClim were above 0.7 for 95% of all species, which compare well with other multispecies comparisons (Elith *et al.*, 2006; Heikkinen *et al.*, 2012). This would lead us to conclude that our models performed well for 95% of species, but in fact only 40% of species had AUCs that were significantly superior to the corresponding null AUCs. This is because the AUC values for distributions generated using our null model, which retains the actual spatial dependence but not the possible association with climate predictors, were centred *above* the expected 0.5 value (see Fig. 2d, h for examples of two null AUC distributions) due to the spatial dependence present within both species distributions and climatic grids.

Our finding that very few rain forest species were significantly related to routinely used climate data reinforces concerns raised by other studies of association for multispecies distribution models in temperate regions (Beale *et al.*, 2008; Chapman, 2010). This weakens the conclusions of SDM studies that do not acknowledge that species distributions may match spatial patterns of climate by chance. While we show that incorporating remote sensing data could substantially improve the modelled association and its transferability for tropical rain forest species, the proportion of climate-dependent species remains low for AB and ACA, whereas most WG species were shown to have a climate-dependent distribution (Fig. 3). Several reasons may be invoked to explain this. First, all WG species were trees, perhaps more directly dependent on the included climatic factors than the different life-forms in the other datasets. However, in ACA, epiphytic Orchidaceae more frequently matched climate than tree species (results not shown). Second, it is likely that non-environmental spatial constraints, known to affect species distributions (Blach-Overgaard *et al.*, 2010), are weaker in WG since it covers a much smaller area than ACA and AB. Third, SDM uncertainties in AB are increased both by the low density and extremely spatially clumped distribution of available collections (Hopkins, 2007; Cayuela *et al.*, 2009) and the inadequate representation of certain habitats in interpolated climatic datasets (Killeen *et al.*, 2007; Buermann *et al.*, 2008). This is in contrast to the more spatially comprehensive species sampling in WG. Finally, the relatively sharp climatic gradients in WG would result in stronger climate filtering.

The slight decrease in model transferability detected when using CRU precipitation data instead of WorldClim (Fig. 3b) confirms the added value of WorldClim over CRU. On the other hand, CRU temperature data allowed for better model fits in WG and AB (Fig. 3). This highlights the non-homogeneous nature of climate data uncertainty, especially for variables like temperature for which there are few ground observations in the tropics (Fig. 1). Our results support the fact that inaccuracies in climate estimates at the 0.1° scale were weaker in MODIS and CHIRPS than in the corresponding WorldClim and CRU data. This is expected to increase at higher sampling rates (Daly, 2006), thus data sampled at even higher resolutions (1 km for WorldClim for instance) should be used with caution.

Remote sensing data captures complex climatic features better

In order to relate the observed variability in modelling performances (Fig. 3) with differences in the climatic patterns of the various datasets, it would be instructive to look more closely at particular topographical features or zones that have either been previously documented or for which we have prior knowledge.

Regarding precipitation, in WorldClim and CRU data we cannot distinguish the rain shadow effect of Mount Cameroon, which is visible on CHIRPS and TRMM (ACA, Fig. S2). Another feature is the size and amplitude of the rainfall gap in the orographic precipitation band over the WG, which is variable between datasets (Fig. S3). The shape, location and number of wet areas on the eastern slopes of the Andes in AB are different for remotely sensed and interpolated datasets (Fig. S4). This discrepancy, among others, has been attributed to data extrapolation from too few ground-based sites in the tropical highland and lowland regions of South America (Killeen *et al.*, 2007). The differences in transferability and species–climate association related to the source of climate datasets are more evident in WG. This 1600-km long mountain ridge extending along the western coast of India is characterized by a sharp precipitation gradient creating a rain shadow effect (Pascal, 1988), both of whioch are difficult to represent by spline surface interpolations as used in CRU and WorldClim (Daly, 2006). The superiority of CHIRPS to the TRMM data is not surprising since they use infrared satellite precipitation estimates including TRMM as input in addition to a large network of ground observations (45,707 weather stations world-wide). Moreover, compared with the 0.05° native resolution of CHIRPS, the 0.25° resolution of TRMM would fail to represent many important smaller-scale climatic features that may influence species distribution.

The discrepancies for annual mean temperature between the datasets are even more conspicuous. Due to paucity of data, interpolation methods often resolve temperature using the elevation gradient alone. In each of the three considered rain forest areas, WorldClim and CRU temperature patterns both closely match the elevation gradient (see Figs S2–S4 for the correlation between estimated climate surface and elevation). This remains partially true for MODIS data, which are nevertheless sensitive to other drivers such as differences in surface properties. For instance, despite their lower elevation, the flooded forest bands around the Amazon (AB), Sangha and Congo (in ACA) rivers appear cooler than neighbouring areas in MODIS, while they are either not distinguished or substantially warmer in WorldClim or CRU. Inversely, the savanna area of the Plateau Batéké (ACA) appears cooler in WorldClim and CRU because of its higher elevation, but is correctly represented as warmer in MODIS. The lowland Beni Savanna (AB) is sharply delineated from the surrounding moist forests as a warmer area in MODIS but is indistinguishable from them in WorldClim and CRU. Finally, the föhn wind effect in WG (Pascal, 1988) and near Mount Cameroon (ACA) is apparent only in the MODIS data with higher temperatures on the eastern slopes, i.e. opposite to what would be expected due to elevation alone.

The ability of space-borne MODIS to better capture the above complex climatic features stems from the intrinsic nature of remote sensing data. Being a direct measure of surface temperatures, independent of weather station observations, remote sensing data have several advantages. As such, they can adequately incorporate important drivers of surface temperature such as slope and aspect on the basis of solar radiation and wind, that are difficult to resolve with standard interpolation (Daly, 2006). Second, although correlated to the temperature recorded 2 m above-ground in meteorological stations (see Fig. 1 for the good correspondence between WorldClim and MODIS in areas with dense station records), surface temperature incorporates information on land cover and water budget. For instance, bare ground and a nearby dense forest will have very different surface temperatures due to differences in albedo, heat capacity and evapotranspiration, which would not be represented by interpolated data unless these are specifically used as predictors. Lastly, the bias of weather stations toward non-vegetated urban areas (Davey & Pielke, 2005), the so-called urban heat island effect, does not affect satellite estimates.

Scope and limitation of remote sensing data

The shorter time coverage of MODIS (2001–13) and TRMM (1998–2013) data compared with WorldClim (1950–2000) might appear problematic. However, in general, remotely sensed climate patterns follow the large-scale spatial gradients of interpolated data with long-term coverage (Figs 1 & S2–S4), the main discrepancies arising in areas with scarce data or complex features which are better captured by remote sensing data. In addition, MODIS temperature data performed better, both in terms of tests of transferability and null model-based species–climate association, than all other temperature data in our study, suggesting that the MODIS time series is already adequate for producing more useful climatologies than interpolated long-term station data.

Because land surface temperature is tightly related to land cover, which is itself highly affected by human activities, incorporation into past and future climate scenarios would present a challenge. However, this limitation does not apply for precipitation data; AFRICLIM already provides projections in future and historical times using, among others, WorldClim, CRU and CHIRPS data as baselines and diverse general circulation and regional climate models (Platts *et al.*, 2015).

Handling and pre-processing remote sensing data before use is often intimidating for modellers. For this study we created a global dataset containing the 19 bioclimatic variables described in Table S1 using MODIS MOD11C3 v.5 and CHIRPS v.2.0 data. This dataset is offered for download at a resolution of 0.10° and 0.05° at https://vdeblauwe.wordpress.com.

Measuring temperature and rainfall from space is a challenge, and sensors and retrieval algorithms are in continuous development to improve the scale and accuracy of records. The end of the TRMM mission approaches and measurements will continue with extended capabilities thanks to the Global Precipitation Measurement (GPM, NASA/JAXA) which started in 2014. Land surface temperature is a key parameter in the physics of land surface processes, and thus receives the attention of a large community of researchers to improve the algorithms for estimating temperature from thermal infrared data (Wan, 2014). CHIRPS data are still in development and improved versions are expected to be released in the coming years (P. Peterson, personal communication).

CONCLUSION

We demonstrate the superiority of remote sensing derived climate data over the widely used interpolation of ground observations to describe actual species climatic filters and to make predictions in novel climatic conditions. Our findings are robust to the choice of modelling algorithm and study region. The detected improvements have important implications for forecasting present, future and past species distributions, prediction of potential invasive ranges and detection of causal climatic drivers of species distribution in the tropics and other station-sparse areas. Our conclusions not only apply to the field of SDM but to any ecological study aimed at capturing the essential climate drivers of species occurrence and diversity in the tropics. We encourage the use of climate data derived from remote sensing rather than interpolated data relying on spatial and topographical predictors (e.g. WorldClim or CRU), at least in areas with sparse weather stations, such as the tropics. This would optimize model predictive performances and modelled association between species distribution and climatic predictors.

ACKNOWLEDGEMENTS

We would like to thank Pete Peterson for useful details about the CHIRPS data and an anonymous referee whose comments helped us to improve the manuscript. This work was partly funded by the French Foundation for Research on Biodiversity (FRB) through its Centre for Synthesis and Analysis of Biodiversity data (CESAB) programme, as part of the RAINBIO research project. V.D. was supported by the IRD post-doctorate fellowship 2013–15.

REFERENCES

Barbet-Massin, M. & Jetz, W. (2014) A 40-year, continent-wide, multispecies assessment of relevant climate predictors for species distribution modelling. *Diversity and Distributions*, **20**, 1285–1295.

Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberon, J. & Villalobos, F. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810–1819.

Beale, C.M., Lennon, J.J. & Gimona, A. (2008) Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences USA*, **105**, 14908–14912.

Bedia, J., Herrera, S. & Gutierrez, J.M. (2013) Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. *Global and Planetary Change*, **107**, 1–12.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**, 289–300.

Blach-Overgaard, A., Svenning, J.C., Dransfield, J., Greve, M. & Balslev, H. (2010) Determinants of palm species distributions across Africa: the relative roles of climate, non-climatic environmental factors, and spatial constraints. *Ecography*, **33**, 380–391.

Boyle, B., Hopkins, N., Lu, Z.Y., Garay, J.A.R., Mozzherin, D., Rees, T., Matasci, N., Narro, M.L., Piel, W.H., McKay, S.J., Lowry, S., Freeland, C., Peet, R.K. & Enquist, B.J. (2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics*, **14**, 16.

Bradley, B.A., Olsson, A.D., Wang, O., Dickson, B.G., Pelech, L., Sesnie, S.E. & Zachmann, L.J. (2012) Species detection vs. habitat suitability: are we biasing habitat suitability models with remotely sensed data? *Ecological Modelling*, **244**, 57–64.

Buermann, W., Saatchi, S., Smith, T.B., Zutta, B.R., Chaves, J.A., Mila, B. & Graham, C.H. (2008) Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography*, **35**, 1160–1176.

Cayuela, L., Golicher, D.J., Newton, A.C., Kolb, M., de Alburquerque, F.S., Arets, E.J.M.M., Alkemade, J.R.M. & Perez, A.M. (2009) Species distribution modeling in the tropics: problems, potentialities, and the role of biological data for effective species conservation. *Tropical Conservation Science*, **2**, 319–352.

Chapman, D.S. (2010) Weak climatic associations among British plant distributions. *Global Ecology and Biogeography*, **19**, 831–841.

Daly, C. (2006) Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*, **26**, 707–721.

Davey, C.A. & Pielke, R.A. (2005) Microclimate exposures of surface-based weather stations – implications for the assessment of long-term temperature trends. *Bulletin of the American Meteorological Society*, **86**, 497–504.

Deblauwe, V., Kennel, P. & Couteron, P. (2012) Testing pairwise association between spatially autocorrelated variables: a new approach using surrogate lattice data. *PLoS ONE*, **7**, e48766.

Dobrowski, S.Z., Thorne, J.H., Greenberg, J.A., Safford, H.D., Mynsberge, A.R., Crimmins, S.M. & Swanson, A.K. (2011) Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. *Ecological Monographs*, **81**, 241–257.

Dormann, C.F. (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129–138.

Dormann, C.F., Purschke, O., Marquez, J.R.G., Lautenbach, S. & Schroder, B. (2008) Components of uncertainty in species distribution analysis: a case study of the great grey shrike. *Ecology*, **89**, 3371–3386.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

FAO (2001) FAOCLIM 2.0 a world-wide agroclimatic database. Environment and Natural Resources Working Paper no. 5. Food and Agriculture Organization of the United Nations, Rome, Italy.

Fernandez, M., Hamilton, H. & Kueppers, L.M. (2013) Characterizing uncertainty in species distribution models derived from interpolated weather station data. *Ecosphere*, **4**, art. 61.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. & Michaelsen, J. (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, **2**, 150066.

Goreaud, F., Loussier, B., Ngo Bieng, M.A. & Allain, R. (2004) Simulating realistic spatial structure for forest stands : a mimetic point process. *Interdisciplinary Spatial Statistics Workshop 2004*, December 2–3, 2004, Paris, France. ENGREF and University of Paris I Panthéon-Sorbonne, Paris, France.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Heikkinen, R.K., Marmion, M. & Luoto, M. (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, **35**, 276–288.

Hijmans, R.J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, **93**, 679–688.

Hijmans, R.J. & Graham, C.H. (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, **12**, 2272–2281.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hopkins, M.J.G. (2007) Modelling the known and unknown plant biodiversity of the Amazon Basin. *Journal of Biogeography*, **34**, 1400–1411.

Killeen, T.J., Douglas, M., Consiglio, T., Jorgensen, P.M. & Mejia, J. (2007) Dry spots and wet spots in the Andean hotspot. *Journal of Biogeography*, **34**, 1357–1373.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J. & Simpson, J. (1998) The Tropical Rainfall Measuring Mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, **15**, 809–817.

Murtaugh, P.A. (2014) In defense of *P* values. *Ecology*, **95**, 611–617.

New, M., Hulme, M. & Jones, P. (1999) Representing twentieth-century space–time climate variability. Part I: development of a 1961–90 mean monthly terrestrial climatology. *Journal of Climate*, **12**, 829–856.

New, M., Lister, D., Hulme, M. & Makin, I. (2002) A high-resolution data set of surface climate over global land areas. *Climate Research*, **21**, 1–25.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P. & Kassem, K.R. (2001) Terrestrial ecoregions of the worlds: a new map of life on Earth. *Bioscience*, **51**, 933–938.

Pascal, J.P. (1988) *Wet evergreen forests of the Western Ghats of India*. Institut Français de Pondichéry, Pondicherry, India.

Pélissier, R. & Goreaud, F. (2015) Ads package for R: a fast unbiased implementation of the K-function family for studying spatial point patterns in irregular-shaped sampling windows. *Journal of Statistical Software*, **63**, 1–18.

Peterson, A.T. (2011a) *Ecological niches and geographic distributions*. Princeton University Press, Princeton, NJ.

Peterson, A.T. (2011b) Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography*, **38**, 817–827.

Peterson, A.T. & Shaw, J. (2003) *Lutzomyia* vectors for cutaneous leishmaniasis in southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *International Journal for Parasitology*, **33**, 919–931.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Platts, P.J., Omeny, P.A. & Marchant, R. (2015) AFRICLIM: high-resolution climate projections for ecological applications in Africa. *African Journal of Ecology*, **53**, 103–108.

Ramesh, B.R. & Pascal, J.P. (1997) *Atlas of endemics of the Western Ghats (India): distribution of tree species in the evergreen and semi-evergreen forests*. Institut Français de Pondichéry, Pondicherry, India.

Randin, C.F., Dirnbock, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.

Rapacciuolo, G., Roy, D.B., Gillings, S., Fox, R., Walker, K. & Purvis, A. (2012) Climatic associations of British species distributions show good transferability in time but low predictive accuracy for range change. *PLoS ONE*, **7**, e40212.

Ripley, B.D. (1976) The second-order analysis of stationary point processes. *Journal of Applied Probability*, **13**, 255–266.

Rödder, D. & Lötters, S. (2010) Explanative power of variables used in species distribution modelling: an issue of general model transferability or niche shift in the invasive greenhouse frog (*Eleutherodactylus planirostris*). *Die Naturwissenschaften*, **97**, 781–796.

Seo, C., Thorne, J.H., Hannah, L. & Thuiller, W. (2009) Scale effects in species distribution models: implications for conservation planning under climate change. *Biology Letters*, **5**, 39–43.

Soria-Auza, R.W., Kessler, M., Bach, K., Barajas-Barbosa, P.M., Lehnert, M., Herzog, S.K. & Bohner, J. (2010) Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. *Ecological Modelling*, **221**, 1221–1229.

Storey, J.D. (2003) The positive false discovery rate: a Bayesian interpretation and the *q*-value. *Annals of Statistics*, **31**, 2013–2035.

Vose, R.S., Schmoyer, R.L., Steurer, P.M., Peterson, T.C., Heim, R., Karl, T.R. & Eischeid, J.K. (1992) *The Global Historical Climatology Network: long-term monthly temperature, precipitation, sea level pressure, and station pressure data*. Oak Ridge National Laboratory, Carbon Dioxide Information Analysis Center, Oak Ridge, TN.

Wan, Z.M. (2014) New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, **140**, 36–45.

Wan, Z.M. & Dozier, J. (1996) A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 892–905.

Washington, R., James, R., Pearce, H., Pokam, W.M. & Moufouma-Okia, W. (2013) Congo Basin rainfall climatology: can we believe the climate models? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**, 20120296.

Xu, T. & Hutchinson, M. (2011) *ANUCLIM version 6.1 user guide*. Australian National University, Fenner School of Environment and Society, Canberra.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Description of the climate dataset used in this study.

Table S1 Details of the derived climatic variables used in this study.

Table S2 Description of all taxa included in the analysis.

Figure S1 Yearly rate of published studies dealing with species distribution models.
Figure S2 Climate and elevation data for Atlantic Central Africa.
Figure S3 Climate and elevation data for the Western Ghats of India.
Figure S4 Climate and elevation data for the Amazon Basin.
Figure S5 Effect of the chosen alpha level on the proportion of species determined as significant with the null model approach.

BIOSKETCH

Vincent Deblauwe is an ecologist who is interested in decoupling the processes underlying patterns in vegetation at multiple scales as well as developing statistical and analytical tools for the study of spatial data in ecology.

Editor: David Currie